

Penerapan Metode Bayesian pada Analisis Regresi untuk Menguji Pengaruh Kepadatan Penduduk dan Aktivitas Ekonomi terhadap Kasus Terkonfirmasi Covid-19 di Indonesia

Meysani Eka Puteri Utomo*, Suliadi

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*puterimeysani05@gmail.com, suliadi@gmail.com

Abstract. In the formation of the regression model, parameter estimation is carried out to produce regression coefficients for each independent variable. One of the methods used to estimate the parameters is the Bayesian test method. The bayesian test method carried out by Lu (et al) (2016) is a bayesian approach method in regression analysis by calculating the Bayes Factor (BF) as a statistical test. The bayesian method carried out by Lu, et al (2016) is the result of the Bayesian factor developed utilizing all the initial (prior) information. In addition, Bayes Factor is calculated using statistics commonly used by frequentists. So that it is easier for frequentists to understand and can use the usual statistical software. This method is also able to measure the strength of evidence that supports alternative hypotheses so that the resulting value is more significant and precise. At the beginning of 2020, there were world health problems caused by a new virus, namely a new type of coronavirus (SARS-CoV-2) whose disease is known as Coronavirus disease 2019 (Covid-19). Covid-19 is spread through close human-to-human contact. Population density is closely related to the spread of Covid-19, where a high population level will lead to a high frequency of interaction between individuals making it easier to spread the virus. Meanwhile, the economic activity described by the 2020 GRDP certainly has a relationship with the Covid-19 situation. Because the impact of Covid-19 occurred in various aspects including economic activity. In this study, the Bayesian method approach will be used in regression analysis to examine the effect of population density and economic activity on Confirmed Cases of Covid-19 in Indonesia. The data used is quantitative data, namely secondary data obtained from the Coordination and Information Center regarding Covid-19 in Indonesia (Covid-19 Distribution Data, 2020) and the Central Statistics Agency. By using the Bayesian method in the simultaneous test, the value $BF_{01} = 5,274391644 \times 10^{-14}$ means that H_1 is accepted and it can be concluded that there is an effect of population density and GRDP 2020 variables on the number of confirmed cases of Covid-19 in Indonesia, with a value of $BF_{10} = 1,895953254 \times 10^{13}$ as proof of acceptance of H_1 in the extreme category, which means that it is very, very strong to accept H_1 because the value of $BF_{10} > 100$.

Keywords: Bayesian Method, Regression Analysis, Population Density, GRDP 2020 and Covid-19.

Abstrak. Dalam pembentukan model regresi dilakukan pendugaan parameter untuk menghasilkan koefisien regresi pada setiap variabel independent. Salah satu metode yang digunakan untuk menduga parameter yaitu metode pengujian Bayesian. Metode pengujian bayesian yang dilakukan oleh Lu (dkk) (2016) merupakan metode pendekatan bayesian pada analisis regresi dengan menghitung Bayes Factor (BF) sebagai uji statistik. Metode bayesian yang dilakukan oleh Lu, dkk (2016) yaitu hasil bayes faktor yang dikembangkan memanfaatkan semua informasi awal (prior). Selain itu, Bayes

Faktor dihitung dengan menggunakan statistik yang biasa digunakan oleh frequentist. Sehingga lebih mudah dipahami oleh frequentist serta dapat menggunakan software statistik yang biasa. Metode ini juga mampu mengukur kekuatan bukti yang mendukung hipotesis alternatif sehingga nilai yang dihasilkan lebih signifikan dan tepat. Awal tahun 2020, telah terjadi masalah kesehatan dunia yang disebabkan oleh virus baru yaitu coronavirus jenis baru (SARS-CoV-2) yang penyakitnya dikenal sebagai *Coronavirus disease 2019* (Covid-19). Covid-19 ini menyebar melalui kontak erat dari manusia ke manusia. Kepadatan penduduk sangat erat kaitannya dengan penyebaran Covid-19, dimana tingkat kependudukan yang tinggi akan menimbulkan tingginya frekuensi interaksi antar individu sehingga lebih mudah untuk menyebarkan virus. Sementara itu aktivitas ekonomi yang digambarkan oleh PDRB 2020 tentunya memiliki hubungan dengan adanya Covid-19 ini. Karena dampak Covid-19 terjadi diberbagai aspek termasuk aktivitas ekonomi. Dalam penelitian ini akan digunakan pendekatan metode bayesian pada analisis regresi untuk menguji pengaruh kepadatan penduduk dan aktivitas ekonomi terhadap Kasus Terkonfirmasi Covid-19 di Indonesia. Data yang digunakan merupakan data kuantitatif yaitu data sekunder yang diperoleh dari Pusat Koordinasi dan Informasi mengenai Covid-19 di Indonesia (Data Sebaran Covid-19, 2020) dan Badan Pusat Statistik. Dengan menggunakan metode bayesian pada uji simultan didapatkan nilai $BF_{01} = 5,274391644 \times 10^{-14}$ yang artinya H_1 diterima dan dapat disimpulkan bahwa terdapat pengaruh variabel kepadatan penduduk dan PDRB 2020 terhadap jumlah kasus terkonfirmasi Covid-19 di Indonesia, dengan nilai $BF_{10} = 1,895953254 \times 10^{13}$ sebagai bukti penerimaan H_1 berada pada kategori *extreme* yang artinya sangat sangat kuat menerima H_1 karena nilai $BF_{10} > 100$.

Kata Kunci: Metode Bayesien, Analisis Regresi, Kepadatan Penduduk, PDRB 2020 dan Covid-19.

1. Pendahuluan

Analisis regresi merupakan salah satu metode statistika yang digunakan untuk mengetahui hubungan antara variabel bebas (*independent*) dan variabel terikat (*dependent*). Menurut (Kutner dkk, 2004) analisis regresi adalah metode statistik yang bertujuan untuk mengetahui hubungan antara dua atau lebih variabel sehingga variabel respon atau hasil dapat diprediksi. Dalam pembentukan model regresi dilakukan pendugaan parameter model regresi yang bertujuan untuk menghasilkan koefisien regresi pada setiap variabel independent. Metode yang sering digunakan untuk menduga parameter yaitu metode kuadrat terkecil (*ordinary least square/OLS*) dan metode kemungkinan maksimum (*maximum likelihood estimation/ML*) yang termasuk ke dalam metode *frequentist*.

Namun metode tersebut menurut Lu, dkk (2016) memiliki kelemahan diantaranya yaitu: pengambilan suatu kesimpulan penelitian bergantung pada aturan “P-value $\leq 0,05$ ” untuk menolak hipotesis nol. Berdasarkan studi yang dilakukan oleh Berger dan Selke (1987 dalam Lu, dkk, 2016), Inferensi menggunakan model regresi frequentis hanya memberikan tingkat kebenaran hipotesis nol sekitar 30% ketika $p=0,05$. Kelemahan kedua metode frequentist yaitu data menjadi satu-satunya sumber informasi, hal tersebut cenderung menghindari penggunaan informasi dari luar. Untuk mengatasi kelemahan pada metode frequentist maka muncul metode bayesian sebagai solusi.

Metode pengujian bayesian merupakan metode yang kegunaanya sama persis seperti metode *frequentis* yaitu sebagai pengujian hipotesis. Namun, perbedaannya dengan *frequentis* yaitu dalam perhitungannya metode bayesian menggunakan bayes faktor sebagai statistik uji yang nantinya mampu mengukur kekuatan bukti yang mendukung hipotesis alternatif. Sehingga dapat dihitung seberapa besar kekuatan menolak atau menerima hipotesis tersebut. Bayes faktor dapat dikatakan sebagai rasio kemungkinan likelihood dari kemungkinan marginal dari dua hipotesis: H_0 dan H_1 , atau dapat dikatakan bahwa faktor bayes merupakan kemungkinan marginal dari dua model (Held & Bove, 2014). Metode bayesian yang dilakukan oleh Lu, dkk

(2016) yaitu hasil bayes faktor yang dikembangkan memanfaatkan semua informasi awal (*prior*). Selain itu, Bayes Faktor dihitung dengan menggunakan statistik yang biasa digunakan oleh *frequentist*. Sehingga lebih mudah dipahami oleh *frequentist* serta dapat menggunakan *software* statistik yang biasa.

Awal tahun 2020, telah terjadi masalah kesehatan dunia yaitu Coronavirus disease 2019 (Covid-19). Penyebaran Covid-19 sangat cepat melalui kontak erat dari manusia ke manusia. Kepadatan penduduk sangat erat kaitannya dengan Covid-19 dimana kepadatan penduduk tinggi akan menimbulkan tingginya frekuensi interaksi antar individu sehingga lebih mudah untuk menyebarkan virus. Sementara itu aktivitas ekonomi yang digambarkan oleh PDRB 2020 tentunya memiliki hubungan dengan adanya Covid-19 ini. Karena dampak Covid-19 terjadi diberbagai aspek termasuk aktivitas ekonomi.

Berdasarkan kasus tersebut, peneliti melakukan penelitian mengenai pengaruh kepadatan penduduk dan aktivitas ekonomi dengan menggunakan pendekatan metode bayesian yang dilakukan oleh Lu, dkk (2016). Maka perumusan masalah dalam penelitian ini yaitu “Bagaimana penerapan metode bayesian pada analisis regresi untuk menguji pengaruh kepadatan penduduk dan aktivitas ekonomi terhadap jumlah kasus terkonfirmasi Covid-19 di Indonesia?”. Tujuan dalam penelitian ini yaitu untuk mengetahui penerapan metode bayesian pada analisis regresi untuk menguji pengaruh kepadatan penduduk dan aktivitas ekonomi terhadap jumlah kasus terkonfirmasi Covid-19 di Indonesia.

2. Metodologi

Bahan

Data yang digunakan dalam penelitian ini merupakan data kuantitatif yaitu data sekunder yang diperoleh dari beberapa sumber resmi. Data tersebut di peroleh dari Pusat Koordinasi dan Informasi mengenai Covid-19 di Indonesia (Data Sebaran Covid-19, 2020) dan Badan Pusat Statistik (BPS) tahun 2020. Komponen dari penelitian ini yaitu berdasarkan 34 provinsi di Indonesia (BPS,2020) dengan skala pengukuran interval.

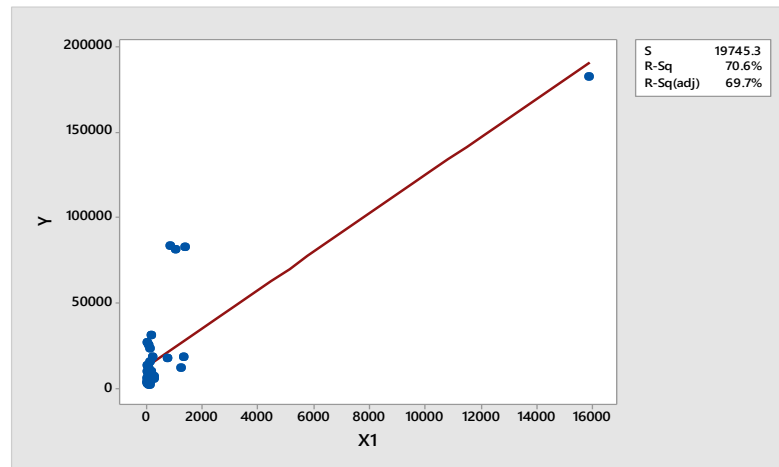
Variabel Penelitian

Variabel yang digunakan dalam penelitian ini disajikan dalam tabel sebagai berikut:

Tabel 1. Definisi Operasional

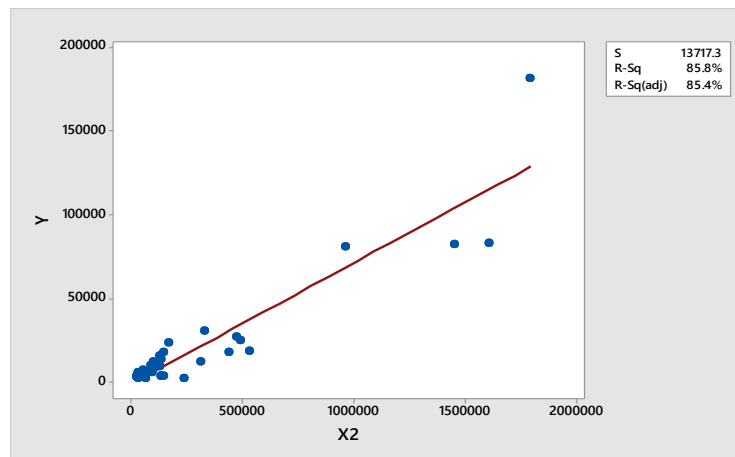
Variabel	Definisi	Satuan
Kepadatan Penduduk (X1)	Menunjukkan banyaknya jumlah penduduk untuk setiap kilometer persegi luas wilayah.	Jiwa/km ²
Produk Domestik Regional Bruto (X2)	Merupakan hasil transaksi yang dilakukan masyarakat berupa jumlah nilai tambah yang dihasilkan dari seluruh kegiatan perekonomian suatu daerah tertentu pada periode tertentu. PDRB yang digunakan merupakan periode tahun 2020.	Milyar/rupiah
Jumlah kasus terkonfirmasi Covid-19 (Y)	Jumlah kasus terkonfirmasi positif Covid-19 pada tahun 2020.	Jiwa

3. Pembahasan dan Diskusi
Deskripsi Data



Gambar 1. Diagram Scatterplot kepadatan penduduk terhadap jumlah kasus terkonfirmasi Covid-19.

Dari hasil plot pada Gambar 1 menunjukkan bahwa variabel kepadatan penduduk (X1) memiliki hubungan dengan jumlah kasus terkonfirmasi Covid-19 (Y) dengan nilai *R-square* sebesar 69,7%.



Gambar 2. Diagram Scatterplot PDRB 2020 terhadap jumlah kasus terkonfirmasi Covid-19.

Dari hasil plot pada Gambar 2 menunjukkan bahwa variabel aktivitas ekonomi yaitu PDRB 2020 (X2) memiliki hubungan dengan jumlah kasus terkonfirmasi Covid-19 (Y) dengan nilai *R-square* sebesar 85,4%.

Uji Asumsi

Uji Multikolinieritas

Berdasarkan hasil perhitungan menggunakan minitab sebagai berikut:

Tabel 2. Hasil Uji Multikolonieritas

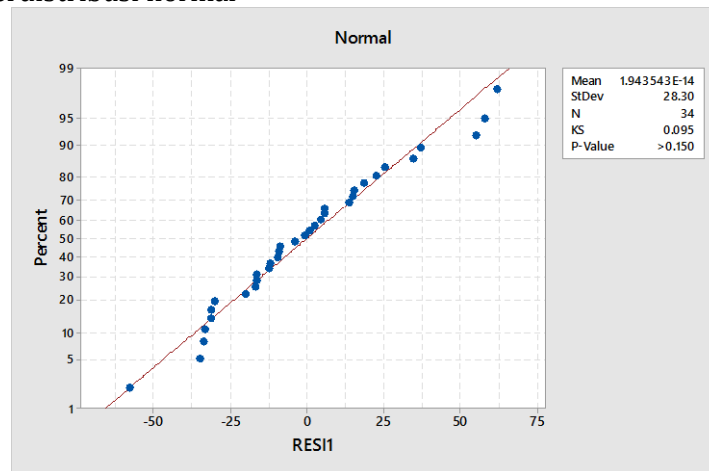
Variabel	VIF
X1	1,69
X2	1,69

Dari tabel 4.2 diperoleh informasi bahwa untuk variabel X1 dan X2 memiliki nilai VIF < 10 maka dinyatakan tidak terdapat gejala multikolonieritas.

Uji Normalitas

H_0 : residual berdistribusi normal

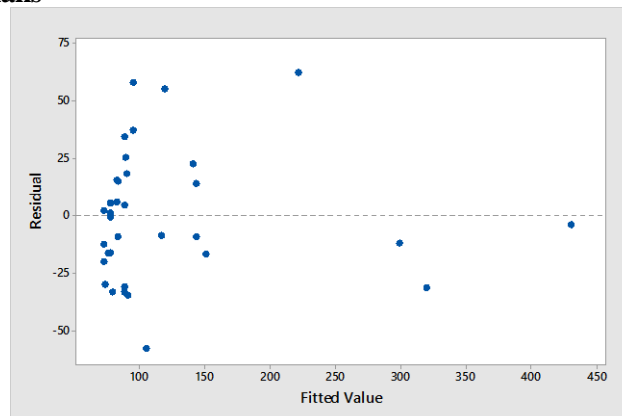
H_1 : residual tidak berdistribusi normal



Gambar 3. Plot Uji Normalitas

Dari output di atas menunjukkan bahwa nilai P-Value dari hasil uji normalitas untuk residual menggunakan Kolmogorov-Smirnov dengan nilai signifikansi sebesar 5% adalah lebih besar dari signifikan ($0,150 > 0,05$), sehingga H_0 diterima. Artinya dapat dikatakan bahwa data residual berdistribusi normal.

Uji Homogenitas Varians



Gambar 4. Plot Uji Homogenitas

Dapat disimpulkan dari Gambar 4 tidak ada gejala heterokedasitas karena plot menyebar di atas dan di bawah sumbu 0 tanpa membentuk pola tertentu. Dapat disimpulkan bahwa data berasal dari populasi yang homogen atau sama.

Model Regresi Linier

Setelah dilakukan pengujian analisis regresi menggunakan minitab diperoleh model regresi dengan peubah bebas yaitu Sqrt (Y) sebagai berikut:

$$\text{Sqrt}(Y) = 68,17 + 0,00547X_1 + 0,000153X_2$$

Penerapan Bayesian pada Analisis Regresi untuk menguji pengaruh kepadatan penduduk dan aktivitas ekonomi terhadap kasus terkonfirmasi Covid-19 di Indonesia

Dilakukan analisis uji simultan dengan metode bayesian, Adapun hipotesis sebagai berikut:

$H_0: \beta_1 = \beta_2 = 0$ Tidak terdapat pengaruh yang signifikan dari seluruh variabel X terhadap variabel Y.

H_1 : minimal ada $\beta_i \neq 0$ dengan $i = 1, 2$ Terdapat pengaruh yang signifikan dari seluruh

variabel X terhadap variabel Y.

Pada pendekatan metode bayesian dalam analisis regresi dilakukan perhitungan Bayes Factor (BF) sebagai statistik uji. Diketahui nilai $F = 125,56$ dengan $q = 2$ dan $v = 31$, dengan menggunakan derajat kebebasan q dan v maka diperoleh nilai $f_{q,v}$ merupakan nilai F statistik biasa yaitu $125,56$. Berikut dilakukan perhitungan BF dengan menggunakan rumus (18) sebagai berikut:

$$BF_{0:1}^{(2)} = 125,56^{2/2} \left[\frac{1 + 2(31)^{-1}}{1 + (31)^{-1}(2)125,56} \right]^{(31+2)/2}$$

$$BF_{0:1}^{(2)} = 5,274391644 \times 10^{-14}$$

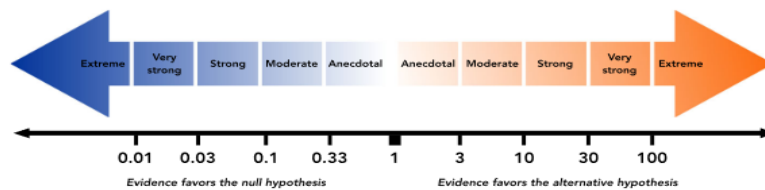
Dari hasil perhitungan di atas di dapat nilai $5,274391644 \times 10^{-14}$.

Dapat disimpulkan bahwa H_0 ditolak dan H_1 diterima karena $5,274391644 \times 10^{-14} < 1$ maka dapat diartikan bahwa terdapat pengaruh yang signifikan dari seluruh variabel X terhadap variabel Y. Kelebihan metode bayesian yaitu mampu melihat kekuatan bukti dalam menerima atau menolak H_0 dengan melakukan perhitungan BF_{10} yang memiliki hubungan dengan BF_{01} , adapun perhitungan sebagai berikut:

$$BF_{0:1}^{(2)} = 5,274391644 \times 10^{-14}$$

$$BF_{10} = \frac{1}{BF_{01}} = \frac{1}{5,274391644 \times 10^{-14}} = 1,895953254 \times 10^{13}$$

Diketahui nilai $BF_{10} = 1,895953254 \times 10^{13}$ dengan menggunakan diagram klasifikasi BF_{10} pada Gambar 5 dapat dikategorikan bahwa bukti penerimaan H_1 berada pada kategori *extreme* yang artinya sangat sangat kuat karena nilai $BF_{10} > 100$.



Gambar 5. Diagram Klasifikasi Penafsiran BF_{10} (Faktor Bayes) (Quintana dan Williams, 2018).

Penerapan Bayesian pada Analisis Regresi untuk menguji pengaruh kepadatan penduduk terhadap kasus terkonfirmasi Covid-19 di Indonesia

Adapun hipotesis untuk kasus ini sebagai berikut:

$H_0: \beta_1 = 0$; Tidak terdapat pengaruh yang signifikan dari variabel kepadatan penduduk (X_1) terhadap jumlah kasus terkonfirmasi Covid-19 di Indonesia (Y).

$H_1: \beta_1 \neq 0$; Terdapat pengaruh yang signifikan dari variabel kepadatan penduduk (X_1) terhadap jumlah kasus terkonfirmasi Covid-19 di Indonesia.

Dengan menggunakan nilai $t = 2,24$; $df = 31$; $n_{cp} = 0$, maka diperoleh nilai BF_{01} menggunakan *software* R sebagai berikut:

$$BF_{01} = 0,0358881$$

Adapun nilai $BF_{01} = 0,0358881$ yang artinya tolak H_0 dan terima H_1 karena nilai $BF_{01} < 1$. Dapat disimpulkan bahwa terdapat pengaruh yang signifikan dari variabel kepadatan penduduk terhadap jumlah kasus terkonfirmasi Covid-19. Selanjutnya, dapat dicari hubungan antara BF_{01} dengan BF_{10} sebagai berikut:

$$BF_{01} = 0,0358881$$

$$BF_{10} = \frac{1}{BF_{01}} = \frac{1}{0,03588881} = 27,86384$$

Dengan menggunakan diagram klasifikasi BF_{10} pada Gambar 5 dapat dikategorikan bahwa bukti penerimaan H_1 berada pada kategori *strong* yang artinya bukti penerimaan H_1 kuat karena nilai BF_{10} berada pada interval $10 \leq BF_{10} \leq 30$.

Penerapan Bayesian pada Analisis Regresi untuk menguji pengaruh aktivitas ekonomi yang ditandai dengan PDRB 2020 terhadap kasus terkonfirmasi Covid-19 di Indonesia

Adapun hipotesis untuk kasus ini sebagai berikut:

$H_0: \beta_2 = 0$; Tidak terdapat pengaruh yang signifikan dari variabel PDRB 2020 (X_2) terhadap jumlah kasus terkonfirmasi Covid-19 di Indonesia (Y).

$H_1: \beta_3 \neq 0$; Terdapat pengaruh yang signifikan dari variabel PDRB 2020 (X_2) terhadap jumlah kasus terkonfirmasi Covid-19 di Indonesia (Y).

Diketahui $t = 10,61$; $df = 31$; $n_{cp} = 0$, dengan menggunakan nilai tersebut maka diperoleh nilai BF_{01} menggunakan *software* R sebagai berikut:

$$BF_{01} = 8,832315 \times 10^{-12}$$

Adapun nilai $BF_{01} = 8,832315 \times 10^{-12}$ yang artinya tolak H_0 dan terima H_1 karena nilai $BF_{01} < 1$. Dapat disimpulkan bahwa terdapat pengaruh yang signifikan dari variabel PDRB 2020 terhadap jumlah kasus Covid-19 di Indonesia. Selanjutnya, dapat dicari hubungan antara BF_{01} dengan BF_{10} sebagai berikut:

$$BF_{01} = 8,832315 \times 10^{-12}$$

$$BF_{10} = \frac{1}{BF_{01}} = \frac{1}{8,832315 \times 10^{-12}} = 1,132205996 \times 10^{11}$$

Dengan menggunakan diagram klasifikasi BF_{10} pada Gambar 5 dapat dikategorikan bahwa bukti penerimaan H_1 berada pada kategori *extreme* yang artinya sangat sangat kuat karena nilai $BF_{10} > 100$.

4. Kesimpulan

Berdasarkan hasil analisis dan pembahasan maka diperoleh kesimpulan sebagai berikut:

1. Berdasarkan hasil pengujian analisis regresi menggunakan bayesian diperoleh model regresi yaitu: $\text{Sqrt}(Y) = 68,17 + 0,00547X_1 + 0,000153X_2$.
2. Berdasarkan hasil pengujian metode bayesian pada uji simultan diperoleh hasil bahwa terdapat pengaruh kepadatan penduduk dan PDRB 2020 terhadap jumlah kasus terkonfirmasi Covid-19 dengan bukti penerimaan H_1 berada pada kategori *extreme* yang artinya sangat sangat kuat menerima H_1 .
3. Berdasarkan hasil pengujian bayesian pada uji parsial untuk kedua variabel sebagai berikut:

Pada pengujian parsial dengan metode bayesian untuk melihat pengaruh kepadatan penduduk terhadap kasus terkonfirmasi Covid-19, diperoleh hasil bahwa terdapat pengaruh kepadatan penduduk terhadap jumlah kasus terkonfirmasi Covid-19 di Indonesia dengan bukti penerimaan H_1 yang kuat (*Strong*). Sementara, pada pengujian parsial dengan metode bayesian untuk melihat pengaruh PDRB 2020 terhadap kasus terkonfirmasi Covid-19 diperoleh hasil bahwa terdapat pengaruh PDRB 2020 terhadap jumlah kasus terkonfirmasi Covid-19 di Indonesia dengan bukti penerimaan H_1 yang sangat-sangat kuat (*Extreme*).

Acknowledge

Penelitian ini dapat terlaksana dengan baik berkat bantuan berbagai pihak, untuk itu peneliti mengucapkan terima kasih kepada Allah SWT, Kedua orang tua dan kepada bapak Suliadi, S.Si., M.Si., Ph.D., yang telah meluangkan waktu untuk membimbing penulis sehingga studi ini selesai. Serta ibu/bapak Dosen Program Studi Statistika Universitas Islam Bandung yang telah membekali penulis ilmu dan teman-teman statistika yang sudah memberikan dukungan dan

semangat hingga penulis dapat menyelesaikan studi ini hingga selesai.

Daftar Pustaka

- [1] BPS. (2020). Produk Domestik Regional Bruto. Retrieved from Badan Pusat Statistik: <https://www.bps.go.id/subject/52/produk-domestik-regional-bruto--lapangan-usaha-.html#subjekViewTab1>
- [2] BPS. (2021, Februari). Ekonomi Indonesia 2020 Turun sebesar 2,07 Persen (c-to-c). Retrieved from Badan Pusat Statistik: [https://www.bps.go.id/pressrelease/2021/02/05/1811/ekonomi-indonesia-2020-turun-sebesar-2-07-persen--c-to-c-.html#:~:text=Perekonomian%20Indonesia%202020%20yang%20diukur,%2Dc\)%20di%20bandingkan%20tahun%202019.](https://www.bps.go.id/pressrelease/2021/02/05/1811/ekonomi-indonesia-2020-turun-sebesar-2-07-persen--c-to-c-.html#:~:text=Perekonomian%20Indonesia%202020%20yang%20diukur,%2Dc)%20di%20bandingkan%20tahun%202019.)
- [3] Data Sebaran Covid-19. (2020). Retrieved from Satuan Tugas Penanganan COVID-19: <https://covid19.go.id/>
- [4] Held, L., & Bove, D. S. (2014). Applied Statistical Inference-Likelihood and Bayes. Springer Verlag Berlin Heidelberg .
- [5] Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). Applied Linear Statistical Models. McGraw-Hill/Irwin.
- [6] Lu, Y., Westfall, P. H., Han, G., & Bui, M. M. (2016). Bayesian Hypothesis Testing for Selected Regression Coefficients. Communications in Statistics - Theory and Methods, 1-29.
- [7] Quintana, D. S., & Williams, D. R. (2018). Bayesian Alternative for Common null-hypothesis significanse tests in psychiatry:a non-technical guide using JASP. BMC Psychiatry.
- [8] Utama Muhammad Bangkit Riksa, Hajarisman Nusar. (2021). *Metode Pemilihan Variabel pada Model Regresi Poisson Menggunakan Metode Nordberg*. Jurnal Riset Statistika, 1(1), 35-42.